Gorjian, Nima and Mittinty, Murthy and Sun, Yong and Yarlagadda, Prasad K. and Ma, Lin (2010) ***Reliability prediction using the non-parametric explicit hazard model : a case study.*** In: 5th World Congress on Engineering Asset Management (WCEAM & AGIC 2010), 25-27 October 2010, Brisbane Convention & Exhibition Centre, Brisbane, Queensland.

RELIABILITY PREDICTION USING THE NON-PARAMETRIC EXPLICIT HAZARD MODEL - A CASE STUDY

Nima Gorjian [a], Murthy N. Mittinty [b], Yong Sun [a], Prasad K.D.V. Yarlagadda [a], Lin Ma [a]

[a] *Cooperative Research Centre for Integrated Engineering Asset Management (CIEAM), School of Engineering Systems, Queensland University of Technology (QUT), Brisbane, Australia*

[b] *School of Mathematical Sciences, Queensland University of Technology (QUT), Brisbane, Australia*

Survival probability prediction using covariate-based hazard approach is a known statistical methodology in engineering asset health management. We have previously reported the semi-parametric Explicit Hazard Model (EHM) which incorporates three types of information: population characteristics; condition indicators; and operating environment indicators for hazard prediction. This model assumes the baseline hazard has the form of the Weibull distribution. To avoid this assumption, this paper presents the non-parametric EHM which is a distribution-free covariate-based hazard model. In this paper, an application of the non-parametric EHM is demonstrated via a case study. In this case study, survival probabilities of a set of resistance elements using the non-parametric EHM are compared with the Weibull proportional hazard model and traditional Weibull model. The results show that the non-parametric EHM can effectively predict asset life using the condition indicator, operating environment indicator, and failure history.

Key Words: Hazard, Reliability, Condition indicator, Operating environment indicator, Explicit hazard model

1 INTRODUCTION

Reliability assessment and health prediction of engineering assets is a significant field of research in engineering asset health management. In reliability analysis, covariate data are often obtained in addition to failure event data. Covariate data are commonly referred to condition data and operating environment data. Condition indicators reflect the degradation level of assets. Some examples of condition indicators are the vibration of fitted rotating machinery, the level of metal particles in engine oil analysis, the sectional loss and wear in a component, to name but a few. On the other hand, operating environment indictors accelerate or decelerate the degradation and failure time of assets. Loads, environmental stresses, and other dynamically changing environment factors are common examples of operating environment indicators. When additional covariate data are available, an alternative approach to classical reliability and survival analysis is the modelling of condition indicators and operating environment indicators via covariate-based hazard models.

Gorjian et al. [1] proposed the Explicit Hazard Model (EHM) to effectively estimate hazards of assets using population characteristics, condition indicators, and operating environment indicators. This model does not have the proportional assumption; as a result, it allows survival curves corresponding to different values of a covariate to cross. The semi-parametric form of this model previously tested using experiment data [2]. The semi-parametric EHM assumes that the baseline hazard follows the Weibull distribution. However, in reality, the analysis of lifetime data of engineering assets often involves complex distributional shapes about which little is known [3]. Therefore, to avoid the restrictive assumption of the semi-parametric EHM, the non-parametric EHM was developed [1]. In this paper, the non-parametric EHM is applied to model the degradation of a set of resistance elements in a case study. The remainder of the paper is organised as follows. Section 2 briefly explains the non-parametric EHM. The parameter estimation of this model is described in Section 3. Survival probabilities of the resistance elements using the non-parametric EHM, Weibull Proportional Hazard Model (WPHM), and traditional Weibull model are shown in Section 4. Conclusions are given in Section 5.

2 NON-PARAMETRIC EXPLICIT HAZARD MODEL (EHM)

The non-parametric EHM is a distribution-free model. It assumes that the baseline hazard is a function of time and condition indicators which provide information about the lifetime of an asset and when it is likely to fail. Operating environment indicators in this model are failure accelerators and/or decelerators that caused by the environment in which an asset operates, and that have not been identified by the condition indicators. The non-parametric EHM is expressed as:

$$h(t;\vec{z}_1(t),\vec{z}_2(t)) = h_0(\exp(\vec{\gamma}_1\vec{z}_1(t)) \, . \, t)\exp(\vec{\gamma}_2\vec{z}_2(t)) \tag{1}$$

Here, $\vec{z}_1(t)$ and $\vec{z}_2(t)$ are vectors of the condition indicator and operating environment indicator, respectively. $\vec{\gamma}_1$ and $\vec{\gamma}_2$ are unknown parameters of the model which define the effects of the condition indicator and operating environment indicator. We follow the study of Shyur et al. [4] to express the baseline hazard of our model. It is assumed that $u = [\exp(\vec{\gamma}_1\vec{z}_1(t)).\, t]$ and u is termed as the baseline time. Suppose a monotone transformation from the baseline time scale u to the observed time scale t is a function of the condition indicator history up to time t. Then, the transformation function is given by:

$$h_0(\exp(\vec{\gamma}_1\vec{z}_1(t)) \, . \, t) = h_0(u) \tag{2}$$

Suppose $u = u[\omega(t), t] = \int_0^t \exp(\vec{\gamma}_1\vec{z}_1(\tau))\, d\tau$, then the non-parametric EHM can be rewritten as:

$$h(t;\vec{z}_1(t),\vec{z}_2(t)) = h_0(u[\omega(t), t])\exp(\vec{\gamma}_2\vec{z}_2(t)) \tag{3}$$

The baseline hazard of Equation (3) needs to be estimated. There are several methods for this approximation [3-6]. A spline function is a natural choice for this approximation where the baseline hazard is affected by a continuous function of the condition indicator [7]. We follow the study of Etezadi-Amoli and Ciampi [3] to estimate the baseline hazard of the non-parametric EHM using a quadratic spline function with one knot. A quadratic spline function with one knot is used in this model to keep the number of parameters small. Another reason to choose this function is that in several cases in litterateur a quadratic spline function with one knot provides a reasonably smooth and accurate fit to data almost as well as a quadratic spline with two knots and a cubic spline with one knot [7, 8]. Suppose λ_j and θ_i for all j and i are coefficients of a spline function, then the quadratic spline with one knot is given by:

$$q(u) = \sum_{j=0}^{2}\lambda_j u^j + \sum_{i=1}^{1}\theta_i\,(u - \xi_i)_+^2 + \epsilon \tag{4}$$

3 PARAMETER ESTIMATION

The key advantage of the non-parametric EHM over the semi-parametric EHM is its parameter estimation. All parameters of this model are estimated without having to make an assumption about the lifetime distribution of the baseline hazard. The partial (or marginal) likelihood function is used to estimate parameters of this model:

$$L(\vec{\gamma}_1,\vec{\gamma}_2,\zeta) = \prod_{i=1}^{n}\frac{h_0(u[\omega_i(t_i), t_i])\exp(\vec{\gamma}_2\vec{z}_{2i}(t_i))}{\sum_{l\in R(t_i)} h_0(u[\omega_l(t_i), t_i])\exp(\vec{\gamma}_2\vec{z}_{2l}(t_i))} \tag{5}$$

Where, n is the number of observed failure times and $R(t_i)$ is the number of items under test at time t_i . The log partial likelihood function is:

$$l(\vec{\gamma}_1,\vec{\gamma}_2,\zeta) = \sum_{i=1}^{n}\left\{\ln\left(h_0(u[\omega_i(t_i), t_i])\exp\left(\vec{\gamma}_2\vec{z}_{2i}(t_i)\right)\right) - \ln\left(\sum_{l\in R(t_i)} h_0(u[\omega_l(t_i), t_i])\exp(\vec{\gamma}_2\vec{z}_{2l}(t_i))\right)\right\} \tag{6}$$

$$l(\vec{\gamma}_1,\vec{\gamma}_2,\zeta) = \sum_{i=1}^{n}\left(\ln h_0(u[\omega_i(t_i), t_i]) + \left(\vec{\gamma}_2\vec{z}_{2i}(t_i)\right)\right) - \sum_{\mathrm{i}=1}^{\mathrm{n}}\ln\left(\sum_{l\in R(t_i)}(h_0(u[\omega_l(t_i), t_i])\exp(\vec{\gamma}_2\vec{z}_{2l}(t_i)))\right) \tag{7}$$

Here, ζ denotes a vector of parameters defining the baseline hazard when a spline function is used. However, a spline function cannot ensure that any point in this function is always positive. It is clear that the hazard must always be greater or equal to zero. Therefore, care should be taken in presenting the baseline hazard using a spline function. Kooperberg et al. [9]

suggested one way of doing that. If $\alpha(t;\vec{z}_1(t)) = \ln h_0(t;\vec{z}_1(t))$, this assumption ensures that the baseline hazard is always positive. Thus, the log partial likelihood function can be rewritten as:

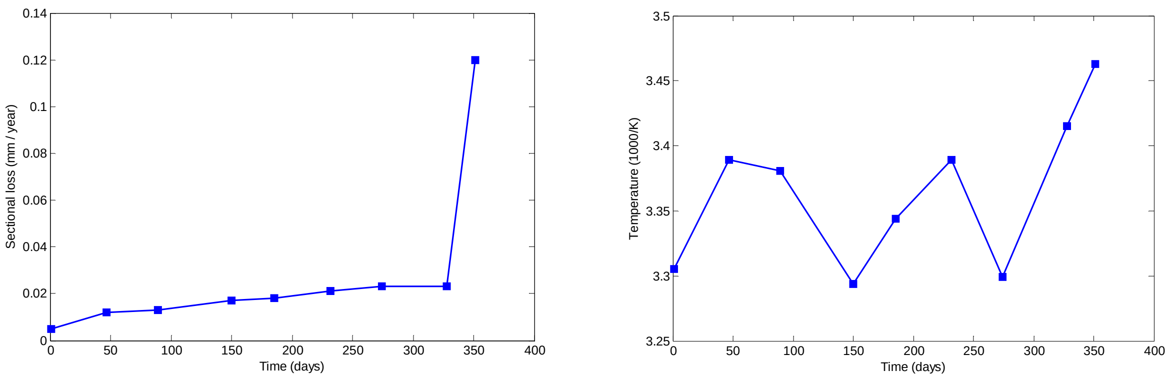
$$l(\vec{\gamma}_1,\vec{\gamma}_2,\zeta) = \sum_{i=1}^{n}\left(\ln h_0(u[\omega_i(t_i),t_i]) + (\vec{\gamma}_2\vec{z}_{2i}(t_i))\right) - \sum_{i=1}^{n}\ln\left(\sum_{l\in R(t_i)}\exp\left(\alpha(u[\omega_l(t_i),t_i])\right)\exp(\vec{\gamma}_2\vec{z}_{2l}(t_i))\right) \quad (8)$$

All parameters are estimated by maximising the log partial likelihood function using a nonlinear optimisation approach.

4 CASE STUDY

Data in this case study were obtained from a laboratory test in the Department of Mechanical Engineering at Monash University (Melbourne, Australia). This laboratory test was conducted using resistance corrosion sensors to measure atmospheric corrosion rates. Condition indicators, operating environment indicators, and failure history of this test were used for this case study (the original test was conducted for other applications). This test was carried out for four resistance elements (i.e. E_1, E_2, E_3, and E_4) on a resistance corrosion sensor board. The typical failure mode of the resistance elements was corrosion. The sectional loss of these resistances and ambient temperature were measured over one year. In this case study, the sectional loss and ambient temperature are considered as the condition indicator and operating environment indicator. As there was no failure event during a year observation, it was assumed that the failure time occurred when the sectional loss reached a pre-specified failure threshold. The trend of data set shows that the sectional loss gradually increased for these resistance elements. The average ranges of changes in sectional loss were between $1.81\mu m$ to $50\mu m$ per year. However, sectional loss values increased beyond $100\mu m$ at certain points in each resistance element. Therefore, it was assumed that failure times occurred at these time points. This type of the failure time is termed as the soft failure where the asset performance deteriorates to an unacceptable level. According to the assumed failure threshold, there were three soft failure times and a suspension time in these resistance elements.

The non-parametric EHM, traditional Weibull model, and WPHM are used in this case study. To avoid overestimating parameters in these models, both the condition indicator and operating environment indicator should be rescaled [10]. The sectional loss rescales from μm to mm. The temperature rescales by $1000/(273+T)$, where T is the temperature in degree Celsius. Figure 1 shows the rescaled values of the sectional loss and ambient temperature for the first resistance element.



(A) **(B)**

Figure 1: (A) Condition indicator and (B) operating environment indicator of the first resistance element (E_1)

The seven parameters of the non-parametric EHM are estimated as: $\hat{\lambda}_0 = 0.0006$, $\hat{\lambda}_1 = 0.0013$, $\hat{\lambda}_2 = 0.0003$, $\hat{\theta}$ =1.0469, $\hat{\xi} = 0.0023$, $\hat{\gamma}_1 = 1.0173$, and $\hat{\gamma}_2 = 0$. $\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}$, and $\hat{\xi}$ are parameters of the spline function. $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are regression coefficients of the condition indicator and operating environment indicator, respectively. Given these parameters, the reliability of the non-parametric EHM can be calculated using the cumulative hazard function $R(t;\vec{z}_1(\tau),\vec{z}_2(\tau)|0\le\tau\le t) = \exp\left[-\int_0^t h(\tau;\vec{z}_1(\tau),\vec{z}_2(\tau))d\tau\right]$.

The shape and scale parameters of the traditional Weibull model are estimated as: $\hat{\beta} = 5.979$ and $\hat{\eta} = 507.986$. The reliability function of this model is given by: $R(t) = \exp\left[-\left(\frac{t}{507.986}\right)^{5.979}\right]$. The shape and scale parameters of the Weibull distribution as well as two regression coefficients of WPHM are estimated as: $\hat{\beta} = 7.5$, $\hat{\eta} = 3828$, $\hat{\gamma}_1 = 7.6$, and $\hat{\gamma}_2 = 3.9$. The reliability function of WPHM is given by: $R(t;\vec{z}(\tau)) = \exp\left[-\int_0^t\left(\frac{\tau}{3828}\right)^{7.5}\exp(7.6\cdot\vec{z}_1(\tau)+3.9\cdot\vec{z}_2(\tau))d\tau\right]$. Here $\vec{z}_1(\tau)$ and $\vec{z}_2(\tau)$ are vectors of the condition indicator and operating environment indictor.

Figure 2 shows the estimated reliability of the individual resistance element (E_1) using the non-parametric EHM and WPHM. It also demonstrates the population reliability of resistance elements using the traditional Weibull model.

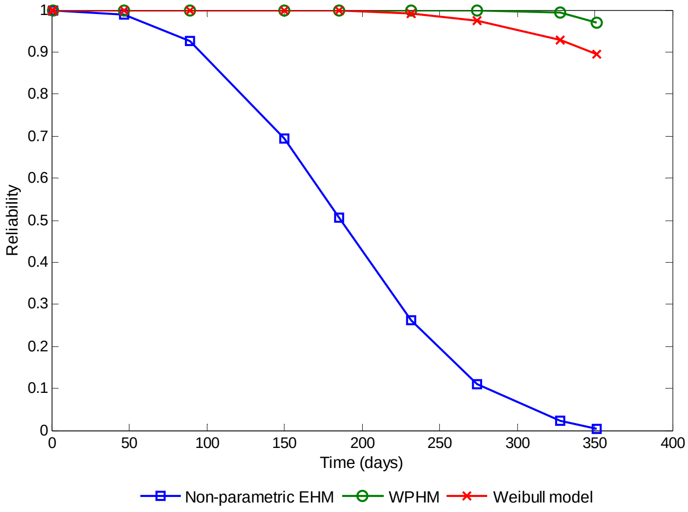


Figure 2: Individual and population reliability

As it can be seen in Figure 1, the sectional loss as a direct condition indicator shows the risk of failure (hazard) better than the operating environment indicator. It seems this direct condition indicator is a more influential indicator in the modelling of asset life. According to the historical failure time data, the first resistance element (E_1) failed on the 351^{th} day. The value of condition indicator significantly increased on the 351^{th} day. On the other hand, the operating environment indicator fluctuated over time and there was no significant change at the 351^{th} day. It could be the reason that the operating environment indicator has no effect in the lifetime modelling of the first resistance element using the non-parametric EHM. In contrast, if the effect of operating environment indicator is zero in WPHM, the estimated survival probability is different to that shown in Figure 2. It is evident that the operating environment indicator has a considerable effect on prediction results of WPHM. However, in this case, the nature of the operating environment indicator shows no significant effect of the data.

This case study demonstrates that while both the condition indicator and the operating environment indicator are included in WPHM; the survival probability of this model is almost similar to that of the traditional Weibull model which does not include any of these indicators in the modelling. Including both of condition and operating environment indicators in WPHM appears to cause a problem in estimation of the regression coefficient and the accuracy of the survival probability prediction. Indeed, the non-parametric EHM shows the better performance in the survival probability estimation where both condition indicators and operating environment indicators are incorporated in the modelling. Moreover, this model does not assume any specified lifetime distribution in the baseline hazard. Therefore, parameters of this model are estimated without having to make an assumption about the lifetime distribution of the baseline hazard.

5 CONCLUSIONS

This paper reports on a case study of the application of the non-parametric Explicit Hazard Model (EHM) to the survival probability estimation of a set of resistance elements. The results of the case study demonstrate that the non-parametric EHM is an appropriate model where the condition indicator and operating environment indicator and their failure-generating mechanisms are used in the modelling. Additionally, it is found that mixing both condition and operating environment indicators in WPHM seems to cause a problem in estimation of the regression coefficient and the accuracy of the survival probability estimation. This research shows that the non-parametric EHM is a distribution free model and becomes promising where the analysis of lifetime data with covariates involves complex distributional shapes about which little is known. This research provides a basis for further studies on the concept of residual life using this model.

6 REFERENCES

1 Gorjian N, Ma L, Mittinty M, Yarlagadda P & Sun Y. (2010) The explicit hazard model - part 1: theoretical development. *Prognostics & System Health Management Conference*, Macau - China. IEEE.

2 Gorjian N, Mittinty M, Ma L, Yarlagadda P & Sun Y. (2010) The explicit hazard model - part 2: applications. *Prognostics & System Health Management Conference*, Macau - China. IEEE.

3 Etezadi-Amoli J & Ciampi A. (1987) Extended Hazard Regression for Censored Survival Data with Covariates: A Spline Approximation for the Baseline Hazard Function. *Biometrics*, 43(1), 181-192.

4 Shyur H-J, Elsayed EA & Luxhøj JT. (1999) A general model for accelerated life testing with time-dependent covariates. *Naval Research Logistics*, 46(3), 303-321.
5 Kalbfleisch JD & Prentice RL. (2002) The statistical analysis of failure time data (Second ed.). New Jersey: Wiley.
6 Whittemore AS & Keller JB. (1986) Survival estimation using splines. *Biometrics*, 42(3), 495-506.
7 Sleeper LA & Harrington DP. (1990) Regression splines in the Cox model with application to covariate effects in liver disease. *American Statistical Association*, 85(412), 941-949.
8 Wold S. (1974) Spline functions in data analysis. *Technometrics*, 16(1), 1-11.
9 Kooperberg C, Stone CJ & Truong YK. (1995) Hazard regression. *American Statistical Association*, 90(429), 78-94.
10 Luxhoj JT & Shyur H-J. (1997) Comparison of proportional hazards models and neural networks for reliability estimation. *Intelligent Manufacturing*, 8(3), 227-234.

Acknowledgement

The authors gratefully acknowledge the financial support provided by both the Cooperative Research Centre for Integrated Engineering Asset Management (CIEAM), established and supported under the Australian Government's Cooperative Research Centres Programme, and the School of Engineering Systems of Queensland University of Technology (QUT).